Estimating a bounded parameter for symmetric distributions

Éric Marchand[a,1,2] , François Perron[b,1] a Département de mathématiques, Université de Sherbrooke, Sherbrooke, Qc, CANADA, J1K 2R1

b Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Qc, CANADA, H3C 3J7

Summary

For the problem of estimating under squared error loss the parameter of a symmetric distribution which is subject to an interval constraint, we develop general theory which provides improvements on various types of inadmissible procedures, such as maximum likelihood procedures. The applications and further developments given include: **(i)** symmetric location families such as the exponential power family including double-exponential and normal, Student and Cauchy, a Logistic type family, and scale mixture of normals in cases where the variance is lower bounded; **(ii)** symmetric exponential families such as those related to a Binomial(n,p) model with bounded $|p-1/2|$ and to a Beta$(\alpha+\theta,\alpha-\theta)$ model; and **(iii)** symmetric location distributions truncated to an interval $(-c,c)$. Finally, several of the dominance results are studied with respect to model departures yielding robustness results, and specific findings are given for scale mixture of normals and truncated distributions.

AMS 2000 subject classifications: 62F10, 62F15, 62F30

Keywords and phrases: Maximum likelihood estimator, restricted parameter space, Bayes estimator, squared error loss, dominance, robustness, symmetric location families, truncated distributions, exponential families, scale mixture of normals.

[1]research supported by NSERC of Canada

[2]Corresponding author; e-mail address: eric.marchand@usherbrooke.ca

1. Introduction

Consider the problem of estimating under squared-error loss, based on an observable X, the location parameter θ of a spherically symmetric univariate model where θ is known to be restricted to an interval $[a, b]$. In their influential 1981 paper, concerned mainly with minimaxity, Casella and Strawderman (1981) showed for $X \sim N(\theta; \sigma^2)$, with known σ^2, that the Bayes estimator associated with the boundary uniform prior $\pi(a) = \pi(b) = 1/2$ is, not only minimax, but dominates as well the maximum likelihood estimator $\widehat{\theta}_{\text{mle}}(X) = XI_{[a \leq X \leq b]} + bI_{[X > b]} + aI_{[X < a]}$ whenever $b - a \leq 2\sigma$.

More recently, Marchand and Perron (2001, 2005) showed, in multivariate versions of the above problem, that the Bayes estimator $\widehat{\theta}_{\text{BU}}$ associated with the uniform prior on $\{\theta : \|\theta - \theta_0\| = m\}$ (known θ_0) dominates $\widehat{\theta}_{\text{mle}}$ on the restricted parameter space $\Theta(m) = \{\theta : \|\theta - \theta_0\| \leq m\}$ whenever $m \leq \sigma\sqrt{p}$ for the models $X \sim \sigma Z + \theta$, with (i) $Z \sim N_p(0, I_p)$, and (ii) $Z \sim$ Multivariate Student with $d \geq p$ degrees of freedom. For the univariate case $p = 1$, their results not only yield Casella and Strawderman's result (with $\theta_0 = (a + b)/2$), but also give an extension to univariate Student distribution with $d \geq 1$ degrees of freedom. Findings concerning the minimaxity of $\widehat{\theta}_{\text{BU}}$ for small enough m were given by Berry (1990), Marchand and Perron (2002), and Marchand and Strawderman (2004).

An interesting question is whether similar results hold for other univariate spherically symmetric models, and to describe as explicitly as possible conditions (on the model and size of the parameter space) for $\widehat{\theta}_{\text{BU}}$ to dominate $\widehat{\theta}_{\text{mle}}$. Applications in this direction do follow from general theory presented in Section 2; where, starting with methods used by Marchand and Perron (2001, 2005), Perron (2003), or Moors (1981, 1985), we obtain elegant and simple extensions (e.g., Theorem 2) applicable to a wide class of spherically symmetric models. But, we actually go beyond in presenting general theory that permits to: **(i)** address situations where the parameter of interest is a function $\eta(\theta)$ (say) of θ, **(ii)** to derive dominance results for more general symmetric models, which include truncated univariate symmetric location distributions and symmetric exponential families, and **(iii)** to handle cases where the target estimator to be dominated is more general and not necessarily $\widehat{\theta}_{\mathrm{mle}}$. Various illustrations are given in Sections 3 and 5. The models studied in Section 3 include Binomial, Beta, the exponential power family including double-exponential and normal, Student and Cauchy, a Logistic type family, and scale mixture of normals in cases where the variance is lower bounded.

As in Marchand and Perron (2005), conditions for which a given estimator $\widehat{\theta}$ dominates $\widehat{\theta}_{\mathrm{mle}}$ on $\Theta(m)$ simultaneously for all members of a class of spherically symmetric distributions may be derived from general theory. In Section 4, we present with the help of results in Section 2 a development in this direction, and we give a key example where the boundary uniform $\widehat{\theta}_{\mathrm{BU}}$ estimator associated with the normal model is shown to dominate simultaneously $\widehat{\theta}_{\mathrm{mle}}$ on $\Theta(m)$ for a large subclass of scale mixture of normals. This last result stands out from previous results of Marchand and Perron (2005) as the dominating $\widehat{\theta}$ possesses nice properties for the model from which it is derived (Bayes, admissible, minimax for small enough m), and at the same time is robust in its dominance of $\widehat{\theta}_{\mathrm{mle}}$ to certain type of departures within the given subclass of scale mixtures.

Our final section (Section 6) is devoted to symmetric location distributions which are truncated to an interval $(-c, c)$. An interesting feature of the general theory resides in the applicability to these types of truncated distributions, which have not been often studied, it seems, in the context of inference in restricted parameter spaces. A further novelty arises as some of the dominance results are robust to the actual value of the above truncation value c (see Remarks 5 and 6), providing further examples of simultaneous dominance. This occurs with the simplicity of the dominance conditions, as well as from the specific property that the estimator $\hat{\theta}_{BU}$ does not depend on the truncation point c.

2. General theory

In this paper, we shall work with symmetric densities $f(\cdot|\theta)$ associated with a symmetric measure μ on $\mathbb{R}$. We will assume that f satisfies the following factorization:

$$f(x|\theta) = \exp -\{h(x-\theta) + \kappa(\theta)\}, \tag{1}$$

for all $x \in \mathbb{R}$, $\theta \in \Theta(m)$ with $\Theta(m) = [-m, m]$, m being fixed. In (1), h is a given continuous and even function. We are concerned here with the estimation of $\eta(\theta)$ under squared error loss $L(\theta, d) = (d - \eta(\theta))^2$, with η being a nondecreasing odd function. Hereafter, we will refer to these components h, κ, μ and η as the components of our general model.

As illustrated with the next examples, many well known families belong to the general model either directly or following re-parametrization.

Example 1. *(Symmetric location families) Let μ be the Lebesgue measure on $\mathbb{R}$. Consider ξ as a location parameter and assume that $\xi \in [a, b]$. Let Y be a continuous random variable and suppose that the density of Y evaluated at y is given by $g(y - \xi)$, g being an even function. By setting $X = Y - (a+b)/2$, $\theta = \xi - (a+b)/2$, $m = (b-a)/2$, and $h = -\log \circ g$, the density of X satisfies factorization (1) with the function κ being constant. Familiar examples of such families include Normal, Laplace, Cauchy and Student, Logistic, exponential power, and scale mixtures of normal distributions.*

Example 2. *(Truncated distributions) Assume that the density of X satisfies factorization criterion (1). Let A be a symmetric set; $A = [-c, c]$ for example. The density of the conditional distribution of X given that $X \in A$ will also satisfy factorization (1) with the same function h, but with κ not constant as above, and the reference measure μ being concentrated on A. The class described above is quite general, but let us nevertheless mention the important case of a truncated normal distribution for which we provide applications in Section 5.*

Example 3. *(Symmetric exponential families) Consider Y a random variable which belongs to an exponential family. Let ν be the reference measure. Assume that ν is symmetric and the density of Y, with respect to ν, depends on a parameter θ and is given by:*

$$\varphi(y|\theta) = \exp\{\theta T(y) - c(\theta)\},$$

where T is an odd function and $\theta \in [-m, m]$. Notice that when $\theta = 0$, Y has a symmetric distribution. Here, if we set $X = T(Y)$, then X has a density satisfying factorization (1) with $h(y) = y^2/2$ for all y, $\kappa(\theta) = c(\theta) - \theta^2/2$ for all $\theta \in [-m, m]$, and

$$\mu(A) = \int_{T^{-1}(A)} \exp(T^2(y)/2)\nu(dy),$$

for any measurable set A. As an application, consider $K \sim$ Binomial(n, p) and set $Y = K - n/2$. Here, ν is the probability measure associated with the random variable $(K - n/2)$ when $K \sim$ Binomial$(n, 1/2)$, T is the identity function, and $\theta = \log(p/(1-p))$.

The general model in invariant under sign changes, and we will hence only consider equivariant estimators, that is estimators which are odd functions. The general model also forces κ to be an even function. This implies that $\hat{\theta}_{\text{mle}}$, the maximum likelihood estimator of θ, is equivariant. A significant portion of our findings concern the comparison of the equivariant estimators $\eta(\hat{\theta}_{\text{mle}})$ (i.e., the mle of $\eta(\theta)$) and $\widehat{\eta(\theta)}_{\text{BU}}$, the Bayes estimator of $\eta(\theta)$ for the uniform prior on the boundary of the parameter space, i.e., $\pi(-m) = \pi(m) = 1/2$. It is easy to verify that

$$\widehat{\eta(\theta)}_{\text{BU}}(x) = \eta(m)\rho_h(m, x), \qquad x \in \mathbb{R}, \tag{2}$$

with

$$\rho_h(\theta, x) = \tanh\{\frac{h(\theta + x) - h(\theta - x)}{2}\}, \qquad \theta, x \in \mathbb{R}. \tag{3}$$

As seen with the following lemma, the function ρ_h plays a key role as well in the risk decomposition of equivariant estimators.

Lemma 1. *Under model (1), and for an equivariant estimator δ of $\eta(\theta)$, we have*

(a) $E_\theta\left[sgn(X)||X|=r\right]=\rho_h(\theta,r)$; $r\geq 0$;

(b) $E_\theta[(\delta(X)-\eta(\theta))^2|\,|X|=r]=[(\delta(r)-\eta(\theta)\rho_h(\theta,r))^2]+\eta^2(\theta)[1-\rho_h^2(\theta,r)]$; $r\geq 0$;

(c) *implying that for $\theta\in\{-\lambda,\lambda\}$; $\lambda\in[0,m]$;*

$$E_\theta[(\delta(X)-\eta(\theta))^2]\geq E_\theta[(\eta(\lambda)\rho_h(\lambda,X)-\eta(\theta))^2];$$

in other words, the equivariant estimator given by $\delta(x)=\eta(\lambda)\rho_h(\lambda,x)$ is best among all equivariant estimators when we know that $\theta\in\{-\lambda,\lambda\}$, and this estimator corresponds to $\widehat{\eta(\theta)}_{BU}$ when $\lambda=m$.

Taken literally, part (c) of Lemma 1 is of limited use because $|\theta|$ is unknown. However, part (b) will be useful as it reveals that the performance of an equivariant estimator δ, taking values $\delta(r)$ for $r>0$, is governed by its proximity to $\eta(\theta)\rho_h(\theta,r)$ as a function of r. Namely, equivariant estimators taking values that are too large in absolute value can be improved upon using Lemma 1. To pursue then, consider $\bar{\rho}_h(m,r)$, the upper envelope of $\rho_h(\theta,r)$ as θ range from 0 to m, i.e,

$$\bar{\rho}_h(m,r)=\sup\{\rho_h(\theta,r);\theta\in[0,m]\},\qquad r\geq 0.$$

Furthermore, for any equivariant estimator δ let

$$A(h,m,\eta,\delta)=\{r\geq 0;\delta(r)>\eta(m)\bar{\rho}_h(m,r)\}.$$

We then have the following results, which are inferred from Lemma 1, and which are similar to Theorems 1 and 2 of Marchand and Perron (2005).

Theorem 1. *Suppose that* $\mu(A(h,m,\eta,\delta_0)) > 0$. *An equivariant estimator* δ *dominates the equivariant estimator* δ_0 *whenever*

$$2\eta(m)\bar{\rho}_h(m,r) - \delta_0(r) < \delta(r) < \delta_0(r)\,, \quad \text{for all}\, r \in A(h,m,\eta,\delta_0),$$

and $\delta(r) = \delta_0(r)$ *for all* $r \notin A(h,m,\eta,\delta_0)$.

Corollary 1. *For the general model with a given* h*, if an equivariant estimator* δ_0 *is such that* $\mu(A(h,m,\eta,\delta_0)) > 0$*, then*

(a) the equivariant estimator δ *with* $\delta(r) = (\eta(m)\bar{\rho}_h(m,r)) \wedge \delta_0(r)$*; for all* $r \geq 0$*; dominates* δ_0*;*

(b) the equivariant estimator δ *with* $\delta(r) = \eta(m)\bar{\rho}_h(m,r)$*; for all* $r \geq 0$*; dominates* δ_0 *whenever* $\eta(m)\bar{\rho}_h(m,r) \leq \delta_0(r)$ *for all* $r \geq 0$*;*

(c) the estimator $\widehat{\eta(\theta)}_{BU}$ *dominates* δ_0 *whenever* $\bar{\rho}_h(m,\cdot) = \rho_h(m,\cdot)$ *and* $\eta(m)\rho_h(m,r) \leq \delta_0(r)$ *for all* $r \geq 0$.

In order to apply Theorem 1 and Corollary 1 we need to know $\bar{\rho}_{h,m}$. For many choices of h and m, the function $\rho_h(\cdot,r)$ will be nondecreasing for all $r \geq 0$, which will imply $\bar{\rho}_h(m,\cdot) = \rho_h(m,\cdot)$. The next lemma provides such key connections linking the behaviour of ρ_h to h and m.

Lemma 2. *(a) The function* $\rho_h(\cdot,r)$ *is nondecreasing on* $(0,m]$ *for all* $r > 0$ *if and only if* h *is nondecreasing on* $(0,\infty)$ *and the expression* $h(\theta+r) - h(\theta-r)$ *is nondecreasing in* θ *on* (r,m) *for all* $r \in (0,m)$.

(b) The function $\rho_h(\theta,\cdot)$ *is nondecreasing on* $(0,\infty)$ *for all* $\theta \in (0,m]$ *if and only if* h *is nondecreasing and convex on* $(0,\infty)$.

Proof. See Appendix.

Remark 1. *The necessary parts of Lemma 2 are not necessary per se, but are presented for sake of completeness and suggest limitations that will arise if one considers loosening conditions on* h*. Part (a) of Lemma 2 depends on the condition saying that* $h(\theta+r)-h(\theta-r)$ *is nondecreasing in* θ *on* (r,m)*, for all* $r\in(0,m)$*. Using arguments similar to those in the proof of part (b) of Lemma 2, this above condition forces* h *to be convex on* $(0,m)$*, while it is satisfied whenever* h *is convex on* $(0,2m)$*. Hence, to summarize, increasing* h *on* $(0,\infty)$ *and convex* h *on* $(0,2m)$ *constitute a simple condition for* $\bar{\rho}_h(m,\cdot)=\rho_h(m,\cdot)$ *to hold, and render part (c) of Corollary 1 applicable.*

We now pursue with a useful property concerning cases where $\eta(\theta)$ is an expectation.

Corollary 2. *Suppose that* T *is a nondecreasing odd function. Let* $\eta(\theta)=E_\theta[T(X)]$*. If* h *is nondecreasing and convex on* $[0,\infty)$*, then* η *is nondecreasing and odd for models in (1).*

Proof. Having verified directly from (1) that η is odd, it is sufficient to show that we have a monotone likelihood ratio to establish the nondecreasing property. Consider $-m\le\theta_0<\theta_1\le m$. Let $\Delta=(\theta_1-\theta_0)/2$ and $y=x-(\theta_1+\theta_0)/2$. We obtain that

$$\frac{f(x|\theta_1)}{f(x|\theta_0)}=\exp\{h(y+\Delta)-h(y-\Delta)+\kappa(\theta_0)-\kappa(\theta_1)\}\,,\ \text{for } y\in\mathbb{R} \text{ and } \Delta\in(0,m].$$

Therefore, for any $\Delta\in(0,m]$, we need to show that the expression $h(y+\Delta)-h(y-\Delta)$ is nondecreasing in y on $\mathbb{R}$. In fact, since h is an even function, we need only to verify that $h(y+\Delta)-h(y-\Delta)$ is nondecreasing in y on $(0,\infty)$. And finally, this last condition follows from the convexity of h.

We will finish this section with a further series of results, consisting of regularity conditions on h and m, which are useful for applying Theorem 1 and Corollary 1. For all $m>0$, consider the functions ϕ and φ_m given by

$$\phi(y)=h(y)-\log(y),\qquad y>0,$$

$$\varphi_m(y)=\{h(m+y)-h(m-y)\}/2,\qquad 0<y<m.$$

For a fixed value of b, $b \in (0, \infty]$, we introduce the following classes of functions:

$$\mathcal{C}_{\mathrm{BU}}(b) = \{h\colon \bar{\rho}_h(m,\cdot) = \rho_h(m,\cdot) \text{ on } (0,\infty), \text{ for all } m \in (0,b]\,\}, \tag{4}$$

$$\mathcal{C}_{\phi}(b) = \{h \in \mathcal{C}_{\mathrm{BU}}(b)\colon \phi'(m+y) + \phi'(m-y) < 2\phi'(m) \text{ for all } y,m \text{ such that } 0<y<m<b\}, \tag{5}$$

$$\mathcal{C}_{\varphi}(b) = \{h \in \mathcal{C}_{\mathrm{BU}}(b)\colon \varphi'_m \text{ is decreasing on } (0,m) \text{ for all } 0<m<b\}. \tag{6}$$

Remark 2. *The characterization of $\mathcal{C}_{BU}(b)$ was previously addressed in Lemma 2 and Remark 1. Assume now that $h \in \mathcal{C}_{BU}(b)$. We have $h \in \mathcal{C}_{\phi}(b)$ implies that ϕ' is concave on $(0,b)$, while ϕ' concave on $(0,2b)$ implies that $h \in \mathcal{C}_{\phi}(b)$. Similarly, $h \in \mathcal{C}_{\varphi}(b)$ implies that h' is concave on $(0,b)$, while h' concave on $(0,2b)$ implies that $h \in \mathcal{C}_{\varphi}(b)$. Hence concave ϕ' and h' further permit quick identifications of members of $\mathcal{C}_{\phi}(b)$ and $\mathcal{C}_{\varphi}(b)$ respectively.*

Lemma 3. *(a) Let δ_0 be an equivariant estimator of $\eta(\theta)$ and assume δ_0 is nonnegative on $(0,\infty)$. We have $\eta(m)\rho_h(m,r) \le \delta_0(r)$ for all $r>0$ if and only if*

$$h(m+r) - \log(\eta(m)+\delta_0(r)) \le h(m-r) - \log(\eta(m)-\delta_0(r)), \tag{7}$$

for all $r>0$ such that $\delta_0(r) < \eta(m)$. In particular, if $h'(m)$ and $\delta_0'(0)$ exist, then we must have $\eta(m)h'(m) \le \delta'(0)$ for (7) to hold.

(b) Let δ_0 be an equivariant estimator of θ with $0 \le \delta_0(r) \le m$ for all $r>0$, and assume η is concave on $(0,m)$. Let $\delta = \eta \circ \delta_0$. If $m\rho_h(m,r) \le \delta_0(r)$ for all $r>0$, then $\eta(m)\rho_h(m,r) \le \delta(r)$ for all $r>0$.

Proof.

(a) The equivalence given by (7) is obtained by working directly with the inequality $\eta(m)\rho_h(m,r) \le \delta_0(r)$ for all $r \ge 0$, while the necessary condition follows by focusing on the neighbourhood of 0 and noticing that

$$\frac{\{h(m+r)-h(m-r)\} - \{\log(\eta(m)+\delta_0(r)) - \log(\eta(m)-\delta_0(r))\}}{2r} \to h'(m) - \delta_0'(0)/\eta(m)\,,$$

as $r \to 0$.

(b) First observe that η is concave with $\eta(0) = 0$ (η is odd), which implies that $\frac{\eta(t)}{t}$ decreases in t; $t > 0$. Exploiting this property along with the assumptions $0 \leq \delta_0(r) \leq m$ and $m\rho_h(m,r) \leq \delta_0(r)$, we have indeed

$$\frac{\eta(m)\rho_h(m,r)}{\eta(\delta_0(r))} \leq \frac{m\rho_h(m,r)}{\delta_0(r)} \leq 1,$$

as was to be shown.

Bringing into play several of the above elements, we now pursue with the application of part (c) of Corollary 1 for the case where the target estimator is given by $\delta_0(r) = m \wedge r$, for all $r \geq 0$, and which corresponds to the maximum likelihood estimator of θ for unimodal symmetric location families of Example 1 (i.e, increasing h).

Theorem 2. *Let $b \in (0, \infty]$ be fixed. If $h \in \mathcal{C}_\phi(b)$ and $bh'(b) \leq 1$, then $m\rho_h(m,r) \leq r$ for all $r \in (0,m)$, $m \in (0,b]$; and consequently $\hat{\theta}_{BU}$ dominates $\hat{\theta}_{mle}$ on $\Theta(m)$; $m \in (0,b]$.*

Proof. Following Lemma 3, it is sufficient to show that $\phi(m+r) \leq \phi(m-r)$ for all $r \in (0,m)$. Since

$$\begin{aligned}
\phi(m+r) - \phi(m-r) &= 2r\phi'(m) + \int_0^r [\{\phi'(m+t) - \phi'(m)\} - \{\phi'(m) - \phi'(m-t)\}]\, dt \\
&\leq 2r\phi'(m) \quad (h \in \mathcal{C}_\phi(b)) \\
&= 2r(h'(m) - 1/m),
\end{aligned}$$

for all $r \in (0,m)$, we have the result.

3. Dominance examples

Example 4. *(Binomial(n,p) with $|p - \frac{1}{2}|$ bounded; Perron, 2003)*

As in the latter part of Example 3, pursue by considering $K \sim$ Binomial(n,p), $Y = K - n/2$, with $\theta =$

$\log(p/(1-p))$; $|\theta| \le m$. *Here we illustrate our dominance results for estimating* $\eta(\theta) = p - 1/2 = \frac{1}{2}\tanh(\theta/2)$ *(equivalent to estimating p), and for extracting conditions under which the estimator* $\widehat{\eta(\theta)}_{BU}$ *dominates* $\widehat{\eta(\theta)}_{mle}$, *where by (2) and (3),* $\widehat{\eta(\theta)}_{BU}(y) = \frac{1}{2}\tanh(\frac{m}{2}) tanh(my)$; *and* $\widehat{\eta(\theta)}_{mle}(y) = \frac{y}{n} \wedge \eta(m); y \ge 0$. *Now, observing that* $\bar{\rho}_h(m,\cdot) = \rho_h(m,\cdot)$ *since* $h(y) = y^2/2 \in \mathcal{C}_{BU}(\infty)$ *(see Remark 1), and applying part (c) of Corollary 1 with* $\delta_0 = \widehat{\eta(\theta)}_{mle}$, *we obtain the sufficient condition for dominance:*

$$\frac{1}{2}\tanh(\frac{m}{2})\tanh(my) \le \frac{y}{n};\ y \ge 0;$$

which occurs whenever $\frac{m}{2}\tanh(\frac{m}{2}) \le \frac{1}{n}$ *(i.e.,* $\frac{\tanh(z)}{z} \le 1$*), or equivalently* $m \le \frac{c}{\sqrt{n}}$ *with* $c \ge 2$.

Example 5. *(A Beta model)*

Consider $Y \sim Beta(\alpha+\theta, \alpha-\theta)$ *with* $|\theta| \le m < \alpha$. *Here is another example of a symmetric exponential family (Example 3) with* $T(y) = \log(y/(1-y))$, *for which we illustrate our dominance results for estimating* θ. *Since* $E(Y) = \frac{1}{2} + \frac{\theta}{2\alpha}$, *a plausible estimator of* θ *is obtained as the truncation* δ_0 *of the unbiased estimator onto* $[-m, m]$, *and given by* $\delta_0(t) = (\alpha\tanh(\frac{t}{2})) \wedge m$, *for* $t \ge 0$. *From (2) and (3), we have* $\widehat{\theta}_{BU}(t) = m\tanh(\frac{t}{2})$; *and, as above in Example 4, applying part (c) of Corollary 1, we obtain the sufficient condition:*

$$m\tanh(mt) \le \alpha\tanh(t/2);\ t \ge 0;$$

for $\widehat{\theta}_{BU}$ *to dominate* δ_0. *Pursuing the analysis, it is seen that dominance holds whenever:* **(i)** $m \le 1/2$; *or* **(ii)** $m > 1/2$ *and* $m \le \sqrt{\frac{\alpha}{2}}$ *which follows with the property* $\frac{\tanh(ax)}{\tanh(bx)} (= \frac{a}{b}\frac{(\tanh(ax))/(ax)}{(\tanh(bx))/(bx)}) \le \frac{a}{b}$ *(as* $\tanh(y)/y$ *decreases for* $y \ge 0$*), which holds for* $x \ge 0$, $a > b$.

The remaining examples deal with unimodal symmetric location families so $\hat{\theta}_{\mathrm{mle}}$ will correspond to $\delta(r) = m \wedge r$ for all $r > 0$. We present applications of part (a) of Theorem 2 to various models investigating the conditions on h and m for which $\hat{\theta}_{\mathrm{BU}}$ dominates $\hat{\theta}_{\mathrm{mle}}$. Basically, given a function h, we wish to identify whether $h \in \mathcal{C}_{\mathrm{BU}}(b)$ (Lemma 2 will be referred to), whether $h \in \mathcal{C}_\phi(b)$, and we want to find the greatest lower bound on b such that $h \in \mathcal{C}_\phi(b)$ and $bh'(b) \le 1$.

Example 6. *(Exponential Power families)*

Consider the exponential power family of distributions (e.g., Box and Tiao, 1973; West, 1987) with densities

$$f(x|\theta) = \frac{\beta}{2\sigma\beta^{1/\beta}\Gamma(1/\beta)} \exp\{-\frac{1}{\beta}(\frac{|x-\theta|}{\sigma})^{\beta}\}, \qquad x \in \mathbb{R}, \tag{8}$$

for parameters $\theta \in \Theta(m)$, $\sigma > 0$, $\beta > 0$ (known). Included in such families are Normal (i.e., $\beta = 2$), the Laplace or Double-Exponential (i.e., $\beta = 1$), and, for $1 \leq \beta \leq 2$, a subclass of scale mixture of normals (see for instance West, 1987). We have

$$h'(y) = \frac{1}{\sigma^{\beta}} y^{\beta-1}, \qquad y > 0.$$

If $\beta \in (0,1)$, then $h \notin \mathcal{C}_{BU}(b)$ for all $b > 0$ as h is not convex (see Remark 1). If $\beta \in [1,2]$, then $h \in \mathcal{C}_{\phi}(\infty)$ as h' is concave on $(0,\infty)$ (see Remark 2). Now, let $c = (2/\{(\beta-1)(\beta-2)\})^{1/\beta}$, $\beta > 2$. For $\beta > 2$, ϕ' is concave on $(0, c\sigma)$ and convex on $(c\sigma, \infty)$. Finally, $bh'(b) \leq 1$ with $b > 0$ if and only if $b \in (0,\sigma]$. Therefore, for $\beta \geq 1$ we have that $\hat{\theta}_{BU}$; which by (2) and (3) is given by $\hat{\theta}_{BU}(x) = m \tanh\{\frac{1}{2\beta\sigma^{\beta}}[(m+|x|)^{\beta} - (|m-|x|)^{\beta}]\}$; dominates $\hat{\theta}_{mle}$ whenever $0 < m \leq \sigma(1 \wedge c/2)$. Notice that $\sigma(1 \wedge c/2)$ may not be the greatest lower bound on b for our sufficient condition of dominance to hold. In fact, Remark 2 tells us that such a lower bound on b belongs to the interval $\big(\sigma(1 \wedge c/2), \sigma(1 \wedge c)\big)$.

Example 7. *Consider families of distributions with*

$$f(x|\,\theta) = \frac{1}{2\sigma B(\alpha,\beta)} \exp\{\alpha|\frac{x-\theta}{\sigma}|\}/(1+\exp\{|\frac{x-\theta}{\sigma}|\})^{\alpha+\beta}, \qquad x \in \mathbb{R} \tag{9}$$

with $\sigma, \alpha, \beta > 0$. These families of distributions may be viewed as a generalization of the Logistic family ($\alpha = \beta = 1$), and include the Hyperbolic Secant location-scale family (i.e., $\alpha = \beta = 1/2$) as well as type III Logistic distributions for $\alpha = \beta$ (e.g., Zelterman and Balakrishnan, 1992). We have

$$h'(y) = \frac{1}{2\sigma}\{(\beta-\alpha) + (\beta+\alpha)\tanh(\frac{y}{2\sigma})\}, \qquad y > 0.$$

The function h is nondecreasing on $(0,\infty)$ if and only if $\beta \geq \alpha$. The function h is convex on $(0,\infty)$. The function ϕ' is concave on $(0,\infty)$. Therefore, if $\beta \geq \alpha$, then $h \in \mathcal{C}_\phi(\infty)$. Finally, $\hat{\theta}_{BU}$ dominates $\hat{\theta}_{mle}$ if $\beta \geq \alpha$ and $m \in (0,b]$, where $b>0$ is such that $bh'(b)=1$.

Example 8. *(Student and Cauchy distributions; Marchand and Perron, 2005) Consider families of distributions in (1) with*

$$f(x|\,\theta) = \frac{1}{B(\nu/2,1/2)\sigma\sqrt{\nu}}\{1+\frac{1}{\nu}(\frac{x-\theta}{\sigma})^2\}^{-\frac{(\nu+1)}{2}},$$

corresponding to Student distributions with positive scale and shape parameters σ and ν. For $\nu=1$, we have the Cauchy distribution. Here, we have

$$h'(y) = (\frac{\nu+1}{\nu})\frac{y}{\sigma^2}\Big(1+\frac{1}{\nu}(\frac{y}{\sigma})^2\Big)^{-1}, \qquad y>0.$$

The function h is nondecreasing on $(0,\infty)$, and convex on $(0,\sigma\sqrt{\nu})$, which tells us that $h \in \mathcal{C}_{BU}(\sigma\sqrt{\nu})$. Moreover, $h \in \mathcal{C}_\phi(\sigma)$ since the function ϕ' is concave on $(0,2\sigma)$. Finally, $bh'(b)=1$ for $b=\sigma$, implying that $\hat{\theta}_{BU}$ dominates $\hat{\theta}_{mle}$ if $m \in (0,\sigma]$.

We consider now the applicability of Theorem 2 to the class of scale mixture of normals, where the distribution of X admits the representation:

$$X|V = v \sim N(\theta, v^{-1}), \tag{10}$$

for some positive random variable V having probability measure τ. Such distributions have densities of the form (1), with constant $\kappa(\theta)$, and

$$h(y) = -\log G(y), \text{ with } G(y) = E[V^{1/2}e^{-y^2V/2}]. \tag{11}$$

Notwithstanding the specific cases treated above (i.e., Logistic in Example 5; Exponential Power families of Example 4 with $1 \leq \beta \leq 2$; Cauchy and Student distributions of Example 8), a general development remains of interest for a fixed h of the form above (for classes of h's where simultaneous dominance occurs, see Section 5). Although, some restrictions on the pair (τ, b) do seem necessary and difficult to specify for the general applicability condition $h \in \mathcal{C}_\phi(b)$ of Theorem 2 to be satisfied, but here are some simple conditions for cases where the mixing parameter V is bounded above by a constant.

Lemma 4. *For scale mixtures as in (10), we have $h \in \mathcal{C}_\phi(b)$ as long as $P[V \leq \frac{1}{4b^2}] = 1$.*

Proof. See Appendix.

Corollary 3. *For scale mixtures as in (10), the condition $P[V \leq \frac{1}{4m^2}] = 1$ is sufficient for $\hat{\theta}_{BU}$ to dominate $\hat{\theta}_{mle}$.*

Proof. Lemma 4 paired with the assumption $P[V \leq \frac{1}{4m^2}] = 1$ tell us that $P[V \leq \frac{1}{4b^2}] = 1$ and $h \in \mathcal{C}_\phi(b)$ for all $b \geq m$. The result now follows from part (a) of Theorem 2 since

$$bh'(b) = b^2 \frac{E[V^{\frac{3}{2}}\, e^{-\frac{b^2 V}{2}}]}{E[V^{\frac{1}{2}}\, e^{-\frac{b^2 V}{2}}]} \leq b^2 (\frac{1}{4b^2}) \leq 1.$$

Remark 3. *The general passage from a single observation to several observations with $(X_1, \ldots, X_n) \sim f_0(x_1 - \theta, \ldots, x_n - \theta)$ is not emphasized in this paper; although it can be handled :* **(a)** *with the sufficient statistic arising with the symmetric exponential families of Example 3,* **(b)** *with the persistence of the scale mixture of normals for $\bar{X}$ whenever $(X_1, \ldots, X_n)|V = v \sim N_n((\theta, \ldots, \theta), vI_n)$ or $X_i|V_i = v_i; i = 1, \ldots, n$; independent $N(\theta, v_i)$, and* **(c)** *by using the conditional distribution of $\bar{X}|T = t$ where $T = (X_1 - X_n, \ldots, X_{n-1} - X_n)$ is the maximal invariant and whenever the distributional assumptions are met.*

4. Examples of simultaneous dominance

Viewing δ_0 and δ as fixed in Theorem 1, the given dominance condition does permit us to investigate whether it is robust to departures in h. An interpretation is that ones starts with a dominance result valid for a fixed model h_0, and obtains the same dominance for certain types of departures h from h_0. Cases of a normal model h_0 are of particular interest. Here, general results are first given (Corollary 4), while applications to scale mixture of normals follow. Further simultaneous dominance results are presented in Section 5.

Corollary 4. *For the general model with a class* H *of* h *functions, an equivariant estimator* δ_0*, and an upper envelope* $V(r) \geq \sup_{h \in H}\{\bar{\rho}_h(m,r)\}$*;* $r \geq 0$*; such that* $\mu\{r \geq 0 : \delta_0(r) > \eta(m)V(r)\} > 0$*,*

(a) the estimator δ *with* $\delta(r) = (\eta(m)V(r)) \wedge \delta_0(r)$ *for all* $r \geq 0$ *dominates* δ_0 *simultaneously for all* $h \in H$*;*

(b) if there exists $h_0 \in H$ *such that* $\bar{\rho}_{h_0}(m,r) = \sup_{h \in H}\{\bar{\rho}_h(m,r)\}$ *for all* $r \geq 0$*, then all three parts of Corollary 1 applied to* h_0 *hold with the dominance occurring simultaneously for all* $h \in H$*;*

(c) moreover as in (b), and for estimating the location parameter θ *of a symmetric location family as in (1), if* $h_0 \in \mathcal{C}_\phi(b)$ *and* $bh'(b) \leq 1$*, then* $\hat{\theta}_{BU}$ *dominates* $\hat{\theta}_{mle}$ *for all* $m \in (0,b]$ *simultaneously for all* $h \in H$*.*

Proof. Part (a) follows from Theorem 1, while part (b) follows from (a) and Corollary 1. Part (c) follows from part (b) and Theorem 2 (a).

A notable feature of the above, and a key element of the proof, resides in the fact that the departures $h \in H$ in part (c) need not be in $\mathcal{C}_\phi(b)$. We pursue with an application of the above Corollary to scale mixtures of normals. The following lemma gives a useful upper envelope for the multiplier $\bar{\rho}_h(m,\cdot)$ in the case of scale mixtures of normals. Moreover, the upper envelope corresponds to a multiplier from a normal distribution, and this is exploited in the corollary subsequent to the following lemma.

Lemma 5. *For scale mixtures of normals as in (10) with* $E(V^{3/2}) < \infty$*, we have* $\bar{\rho}_h(m,r) \leq \tanh(mr\frac{E(V^{3/2})}{E(V^{1/2})})$ *for all* (m,r)*.*

Proof. From Marchand and Perron (2005), or Marchand (1993), we may write $\rho_h(\lambda, r) = E[\tanh(tW)]$ with $t = \lambda r$, $s = (\lambda^2 + r^2)/2$, and W having density (with respect to τ) proportional to $w^{1/2}\cosh(tw)e^{-sw}$. Now by virtue of the concavity of $\tanh(y)$; $y > 0$; and Jensen's inequality, it follows that $\rho_h(\lambda, r) \le \tanh(tE(W))$ with

$$E[W] = \frac{\int_0^\infty w^{3/2}\cosh(tw)e^{-sw}d\tau(w)}{\int_0^\infty w^{1/2}\cosh(tw)e^{-sw}d\tau(w)} = \frac{E[Z\cosh(tZ)e^{-sZ}]}{E[\cosh(tZ)e^{-sZ}]}\,;$$

Z having density (with respect to τ) equal to $r(z) = \frac{z^{1/2}}{E(V^{1/2})}$. Now, by observing that the function $b(z) = \cosh(tz)e^{-sz}$ is decreasing in z, and that consequently the covariance between Z and $b(Z)$ is negative, we infer that $\rho_h(\lambda, r) \le \tanh(tE(Z)) = \tanh(\lambda\, r\,\frac{E(V^{3/2})}{E(V^{1/2})})$ which leads to the desired result as $\tanh(y)$ is increases in y.

Corollary 5. *For the subclass H of scale mixtures in (10) with $\frac{E(V^{1/2})}{E(V^{3/2})} \ge \ell$; which includes all $N(\theta, \sigma^2)$ distributions with $\sigma^2 \ge \ell$; the $\hat{\theta}_{BU}$ estimator associated with a $N(\theta, \sigma^2 = \ell)$ distribution (see Example 6 with $\beta = 2$), given by $\hat{\theta}_{BU}(x) = m\tanh(mx/\ell)$, dominates simultaneously $\hat{\theta}_{mle}$ for all $h \in H$, whenever $m \le \sqrt{\ell}$. On the other hand, if $m > \sqrt{\ell}$, then δ_g with $\delta(r) = (m\tanh(mr/\ell)) \wedge r$ for $r \ge 0$ dominates simultaneously $\hat{\theta}_{mle}$ for all $h \in H$.*

Proof. The result follows directly from Lemma 5 and Corollary 4.

Remark 4. *Marchand and Perron (2005) obtain a similar but weaker simultaneous dominance result to cases where V is bounded above with probability one. On the other hand, their result is set, and applicable in a multidimensional context where the mean θ is upper bounded in norm.*

Example 9. *Consider the subclass of families of Student distributions with degrees of freedom $d \ge d_0$, with $d_0 \ge 1$. Since, Student distributions are scale mixtures of the form (10) with $V \sim Gamma(d/2, d/2)$, it is easy to obtain that $\frac{E(V^{1/2})}{E(V^{3/2})} = d/(d+1)$. Hence, Corollary 5 applies with $\ell = d_0/(1+d_0)$. As an example, take $d_0 = 1$ and $\ell = 1/(1+1) = 1/2$. Corollary 5 tells us that the estimator $m\tanh(2mX)$ dominates*

$\hat{\theta}_{mle}$ *for all* $N(\theta, \sigma^2)$ *models with* $\sigma^2 \geq 1/2$*, as well as for Student distribution models:* $X - \theta \sim T_d$*, and* $X - \theta \sim cT_d$*;* $c > 1$*; with* $d \geq 1$ *degrees of freedom under the constraint* $|\theta| \leq m$ *with* $m \leq \frac{\sqrt{2}}{2}$*.*

5. Truncated continuous distributions

In this section, we focus on the distributions of Example 2 with densities of the form:

$$f(x|\theta) = \exp -\{h(x-\theta) + \kappa_c(\theta)\}\, I_{[-c,c]}(x), \qquad x \in \mathbb{R}, \tag{12}$$

for $\theta \in \Theta(m)$, $c \in (0, \infty]$, with h even and nondecreasing on $(0, \infty)$.

The type of applications obtained here, which seem quite new, arise in part from the simplicity of the dominance conditions, as well as from the specific property that the estimator $\hat{\theta}_{BU}$ does not depend on the truncation point c.

Remark 5. *If the conditions of Theorem 1 or Corollary 1 are satisfied for a model as in (12) with* $c = \infty$*, then dominance results arising from Theorem 1 or Corollary 1 persist for the same model for all* $c > 0$*, as long as* $\mu(A(h, m, \eta, \delta_0) \cap [0, c]) > 0$*. Moreover, observe that the Bayes estimator* $\widehat{\eta(\theta)}_{BU}$ *for a model as in (12) does not depend on c; and matches* $\widehat{\eta(\theta)}_{BU}$ *for the same model with* $c = \infty$*. This signifies for instance that cases of dominance of* $\widehat{\theta}_{mle}$ *by* $\widehat{\theta}_{BU}$*, such as those arising with applications of Theorem 2 or those illustrated in Section 3, hold simultaneously not only for* $c = \infty$*, but also as well for all* $c > 0$*.*

Note that this last remark does not relate to the mle of θ, which depends on c and which we will denote $\hat{\theta}_{\text{mle},c}$. However, the property that the estimator $\widehat{\eta(\theta)}_{BU}$ does not depend on c, along with the ordering of maximum likelihood and $\hat{\theta}_0$ estimates, with $\hat{\theta}_0(r) = m \wedge r$, $r \geq 0$, established with the next lemma, we obtain straightforward extensions to all truncated problems in (12). Along with Remark 5 (and also

Remark 6), this adds considerably to the catalog and richness of applications of our dominance findings, and to the attractiveness of the estimator $\widehat{\eta(\theta)}_{BU}$ for cases where m is sufficiently small.

Lemma 6. *For the model given in (12) we have $\hat{\theta}_{mle,c}(x) \geq \hat{\theta}_0(x)$ for all $x, c > 0$.*

Proof. Assume that $x > 0$ is fixed. Let $z(\theta) = h(x-\theta) + \kappa_c(\theta)$, $\theta \in \Theta(m)$. It is easy to see that $z(|\theta|) \leq z(\theta)$ for all θ. Let us show that $z(x) \leq z(\theta)$ for all $\theta \in [0, x]$. The function κ_c is even and

$$\frac{d}{d\theta} \log \kappa_c(\theta) = \exp\{-h(\theta + c)\} - \exp\{-h(\theta - c)\} < 0$$

for all $\theta > 0$. Moreover, $h(x - \theta)$ is nonincreasing in θ on $[0, x]$. Therefore, the function z is nonincreasing on $[0, x]$, hence the result.

Corollary 6. *If $h \in \mathcal{C}_\phi(b)$, $bh'(b) \leq 1$ and η is concave on $(0, \infty)$, then:* **(i)** *$\widehat{\eta(\theta)}_{BU}$ dominates $\eta(\hat{\theta}_0)$, and* **(ii)** *$\eta(\hat{\theta}_0)$ dominates $\eta(\hat{\theta}_{mle,c})$ for all $c > 0$.*

Proof. It follows from Theorem 2 and Remark 5 that $\widehat{\theta}_{\mathrm{BU}}$ dominates $\hat{\theta}_0$ for the estimation of θ, and for all $c > 0$. As well, from Theorem 2, Theorem 1, and Lemma 6, we obtain that $\hat{\theta}_0$ dominates $\hat{\theta}_{\mathrm{mle},c}$ for the estimation of θ, for all $c > 0$. Finally, Theorem 1 and Lemma 3 lead to the result for the estimation of $\eta(\theta)$.

Remark 6. *Lemma 6 and Corollary 6 apply as well to cases where c is unknown (and where $\eta(\theta)$ does not depend on c). In such cases, the maximum likelihhood estimator of $\eta(\theta)$ is given by $\eta(\hat{\theta}_{mle,}\hat{c}_{mle})$ where $(\hat{c}_{mle}, \hat{\theta}_{mle})$ is the maximum likelihood estimator of (c, θ). This can be proven along the same lines as above, with the additional observation that the likelihood function is maximized in c, for all $\theta \in \Theta(m)$, by $\hat{c}_{mle}(X) = |X|$. Hence, we have, with the results above, examples of the notable feature of a dominance result which persists in models where the truncation point c is unknown.*

We now pursue with further analysis relative to cases where the estimand is $\eta_c(\theta) = \mathrm{E}_{\theta,c}[X]$, and conclude with the specific case of a truncated normal model.

Lemma 7. *We have*

(a) $\eta_c(\theta) < \theta$ *for all* $c, \theta > 0$,

(b) if h *is nondecreasing and convex on* $(0, \infty)$, *then* $\eta_c(\theta)$ *is nondecreasing in* c *on* $(0, \infty)$ *for all* $\theta > 0$;

(c) if h *is increasing on* $(0, \infty)$, *then* $E_\theta[(X - \eta_c(\theta))^3] < 0$ *for all* $c \in (0, \infty)$;

(d) if $h(x) = x^2/(2\sigma^2)$, $x \in \mathbb{R}$, $\sigma^2 > 0$, *then* η_c *is concave on* $(0, \infty)$ *for all* $c > 0$.

Proof.

(a) Using the fact that h is even, we have

$$\eta_c(\theta) - \theta = -\int_{|c-\theta|}^{c+\theta} u \exp\{-h(u) - \kappa_c(\theta)\}\, du < 0$$

for all $c, \theta > 0$.

(b) Let $\theta > 0$, $c \in (0, \infty]$ be fixed, and $R = |X|$. The density of R is given by $f_{\theta,c}$, with

$$f_{\theta,c}(r) = \exp\{-\kappa_c(\theta)\}[exp\{-h(r-\theta)\} + exp\{-h(r+\theta)\}], \qquad 0 \le r < c.$$

For all $0 < c_0 < c_1 \le \infty$, we obtain that

$$\eta_{c_0}(\theta) - \eta_{c_1}(\theta) = \exp\{\kappa_{c_1}(\theta) - \kappa_{c_0}(\theta)\}\mathrm{Cov}_{\theta,c_1}(I_{[0,c_0)}(R), R\rho_h(\theta, R)) < 0\,,$$

because the indicator function $I_{[0,c_0)}$ is nonincreasing on $[0, c_1)$, and the function $\rho_h(\theta, \cdot)$ is nondecreasing on $[0, c_1)$ by virtue of Lemma 2.

(c,d) See Appendix.

Example 10. *(truncated normal) Let* $h(x) = x^2/(2\sigma^2)$, $x \in \mathbb{R}$, $\sigma^2 > 0$. *Let* c *be known. We have* $h \in \mathcal{C}_\phi(\sigma)$, $\sigma h'(\sigma) = 1$, *and* η_c *is concave on* $(0, \infty)$. *We can apply Corollary 6. Therefore, for the estimation of* $\eta_c(\theta)$, *the estimator* $\widehat{\eta(\theta)}_{BU}$ *dominates* $\eta(\hat{\theta}_0)$, *for all* $m \in (0, \sigma]$.

Acknowledgements

The authors thank Idir Ouassou and Amir Payandeh for useful comments. The support of NSERC of Canada is gratefully acknowledged by both authors.

References

Berry, C. (1990). Minimax estimation of a bounded normal mean vector, *Journal of Multivariate Analysis*, **35**, 130-139.

Box, G.E.P. & Tiao, G.C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, Massachusetts: Addison-Wesley.

Casella, G. & Strawderman, W.E. (1981). Estimating a bounded normal mean. *Annals of Statistics*, **9**, 870-878.

Marchand, É. (1993). Estimation of a multivariate mean with constraints on the norm. *Canadian Journal of Statistics*, **21**, 359-366.

Marchand, É. & Strawderman, W. E. (2004). Estimation in restricted parameter spaces: A review. *Festschrift for Herman Rubin*, IMS Lecture Notes-Monograph Series, **45**, pp. 21-44.

Marchand, É. & Perron, F. (2001). Improving on the MLE of a bounded normal mean. *Annals of Statistics*, **29**, 1066-1081.

Marchand, É. & Perron, F. (2002). On the minimax estimator of a bounded normal mean. *Statistics & Probability Letters*, **58**, 327-333.

Marchand, É. & Perron, F. (2005). Improving on the MLE of a bounded mean for spherical distributions. *Journal of Multivariate Analysis*, **92**, 227-238.

Marchand, É., Perron, F., and Gueye, R. (2005) Minimax estimation of a constrained Binomial proportion p when $|p - \frac{1}{2}|$ is small. *Sankhyā: The Indian Journal of Statistics*, **67**, 526-537.

Moors, J.J.A. (1981). Inadmissibility of linearly invariant estimators in the truncated parameter spaces, *Journal of the American Statistical Association*, **76**, 910-915.

Moors, J.J.A. (1985). Estimation in truncated parameter spaces. Ph.D. thesis, Tilburg University.

Perron, F. (2003). Improving on the mle of p for a Binomial(n, p) when p is around 1/2. *Festschrift for Constance van Eeden*, IMS Lecture Notes-Monograph Series, **42**, pp. 45-64.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, **74**, 646-648.

Zelterman, P. & Balakrishnan, N. (1992). Univariate generalized distributions, in *Handbook of the Logistic Distribution*, edited by N. Balakrishnan, Marcel Dekker, New York, 209-221.

6. Appendix

6.1. Proof of Lemma 2

Part (a) The function $\rho_h(\cdot, r)$ is nondecreasing on $[0, m]$ for all $r \geq 0$ if and only if $h(|\theta + r|) - h(|\theta - r|)$ is nondecreasing in θ on $[0, m]$ for all $r \geq 0$.

Suppose that $h(|\theta+r|)-h(|\theta-r|)$ is nondecreasing in θ on $[0, m]$ for all $r \geq 0$. If $0 \leq a < b$ and $b-a \leq 2m$, then

$$\begin{aligned} h(b)-h(a) &= h(|\frac{b-a}{2}+\frac{b+a}{2}|)-h(|\frac{b-a}{2}-\frac{b+a}{2}|) \\ &\geq h(|\theta+\frac{a+b}{2}|)-h(|\theta-\frac{a+b}{2}|) \qquad \text{for } 0<\theta<\frac{b-a}{2} \\ &\to 0 \qquad \text{as } \theta\to 0\,; \end{aligned}$$

so that h is nondecreasing on $[0,\infty)$. If $0<r<\theta<m$, then $h(\theta+r)-h(\theta-r)=h(|\theta+r|)-h(|\theta-r|)$ implying that $h(\theta+r)-h(\theta-r)$ is nondecreasing in θ on (r,m) for all $r\in(0,m)$.

Suppose that h is nondecreasing on $(0,\infty)$ and $h(\theta+r)-h(\theta-r)$ is nondecreasing in θ on (r,m) for all $r\in(0,m)$. If $\theta\leq r$, then $h(|\theta+r|)-h(|\theta-r|)=h(r+\theta)-h(r-\theta)$ and the nondecreasing property of h implies that $h(|\theta+r|)-h(|\theta-r|)$ is nondecreasing in θ on $(0,r]$ for all $r>0$. Finally, if $0<r<\theta<m$, then $h(|\theta+r|)-h(|\theta-r|)=h(\theta+r)-h(\theta-r)$, and the fact that $h(\theta+r)-h(\theta-r)$ is nondecreasing in θ on (r,m) for all $r\in(0,m)$ implies that $h(|\theta+r|)-h(|\theta-r|)$ is nondecreasing in θ on (r,m) for all $r\in(0,m)$.

Part (b) The function $\rho_h(\theta,\cdot)$ is nondecreasing on $[0,\infty)$ for all $\theta\in[0,m]$ if and only if $h(|\theta+r|)-h(|\theta-r|)$ is nondecreasing in r on $[0,\infty]$ for all $\theta\in(0,m]$.

Suppose that $h(|\theta+r|)-h(|\theta-r|)$ is nondecreasing in r on $[0,\infty)$ for all $\theta\in(0,m]$. If $0\leq a<b$ and $a+b\leq 2m$, then

$$\begin{aligned} h(b)-h(a) &= h(|\frac{b+a}{2}+\frac{b-a}{2}|)-h(|\frac{b+a}{2}-\frac{b-a}{2}|) \\ &\geq h(|\frac{b+a}{2}+r|)-h(|\frac{b+a}{2}-r|) \qquad \text{for } 0<r\leq\frac{b-a}{2} \\ &\to 0 \qquad \text{as } r\to 0\,; \end{aligned}$$

so that h is nondecreasing on $(0,m]$ and $h(2\theta)\geq h(0)$ for all $\theta\in(0,m]$. Assuming now that $m<a<b$

and $b-a<2m$, we obtain that

$$\begin{aligned} h(b)-h(a) &= h(|\frac{b-a}{2}+\frac{b+a}{2}|)-h(|\frac{b-a}{2}-\frac{b+a}{2}|) \\ &\geq h(|b-a|)-h(0) \qquad \text{because } \frac{b+a}{2}\geq\frac{b-a}{2} \\ &\geq 0 \qquad \text{because } b-a\in(0,2m]\,, \end{aligned}$$

which tells us that h is nondecreasing on (m,∞). If $0<a<b$ and $b-a<4m$, then we set $\theta=(b-a)/4$, $r_1=a+\theta$ and $r_2=b-\theta$. We obtain that $0<\theta<m$, $0<r_1<r_2$, and

$$\begin{aligned} h(a)+h(b)-2h(\frac{a+b}{2}) &= [h(|\theta+r_2|)-h(|\theta-r_2|)]-[h(|\theta+r_1|)-h(|\theta-r_1|)] \\ &\geq 0. \end{aligned}$$

Since h is continuous, we conclude that h is convex on $(0,\infty)$.

Finally, suppose h is nondecreasing and convex on $(0,\infty)$. If $r\in[0,\theta]$, then $h(|\theta+r|)-h(|\theta-r|)=h(\theta+r)-h(\theta-r)$ so that $h(|\theta+r|)-h(|\theta-r|)$ is nondecreasing in r on $[0,\theta]$ for all $\theta\in[0,m]$. If $r\in(\theta,\infty)$, then $h(|\theta+r|)-h(|\theta-r|)=h(r+\theta)-h(r-\theta)$, and the convexity of h implies that this expression is nondecreasing in r implying that $h(|\theta+r|)-h(|\theta-r|)$ is nondecreasing in r on (θ,∞) for all $\theta\in[0,m]$.

6.2. Proof of Lemma 4

Since, it is clear from (11) that $h(y)$ is increasing for positive y (without conditions on V), it will suffice to establish the stronger conditions (see Remarks 1 and 2) that $h(y)$ is convex and $\phi'(y)=h'(y)-\frac{1}{y}$ is concave for $y\in(0,2b)$, with the given boundedness condition on V. By differentiation, we have from (11):

$$h''(y)=\frac{G'(y)^2-G(y)G''(y)}{G^2(y)},$$

and

$$\frac{\partial^2}{\partial^2 y}\,\{\,h'(y)-\frac{1}{y}\,\} = A(y)+2B(y),$$

with $A(y)=\frac{G'(y)G"(y)-G(y)G'''(y)}{G^2(y)}$ and $B(y)=-\frac{G'(y)h''(y)}{G(y)}-\frac{1}{y^3}$.

Observe first that $G''(y)=E[(y^2V-1)\,V^{\frac{3}{2}}\,e^{-\frac{y^2V}{2}}]$, so that the condition $P[y^2V\le 1]=1$ for all $y\in(0,2b)$; i.e., $P[V\le\frac{1}{4b^2}]=1$; implies the desired convexity.

We pursue by showing that both $A(y)$ and $B(y)$ are non positive for $y\in(0,2b)$ under the given conditions. Turning our attention to $A(y)$, we have

$$\begin{aligned}
\frac{A(y)}{yG(y)} &= \frac{1}{y}[G'(y)G"(y)-G(y)G'''(y)]\\
&= (E[\,V^{\frac{3}{2}}\,e^{-\frac{y^2V}{2}}]E[(1-y^2V)\,V^{\frac{3}{2}}\,e^{-\frac{y^2V}{2}}])-(E[\,V^{\frac{1}{2}}\,e^{-\frac{y^2V}{2}}]E[(3-y^2V)\,V^{\frac{5}{2}}\,e^{-\frac{y^2V}{2}}])\\
&\le (E[\,V^{\frac{3}{2}}\,e^{-\frac{y^2V}{2}}]E[(1-y^2V)\,V^{\frac{3}{2}}\,e^{-\frac{y^2V}{2}}])-(E[\,V^{\frac{1}{2}}\,e^{-\frac{y^2V}{2}}]E[(1-y^2V)\,V^{\frac{5}{2}}\,e^{-\frac{y^2V}{2}}]);
\end{aligned}$$

so that a sufficient condition for $A(y)$ to be non positive is

$$E_{\epsilon=1/2}[(1-y^2W)W\,]\le E_{\epsilon=3/2}[(1-y^2W)W\,]\,,$$

with W having density proportional to $w^{\epsilon}e^{-y^2w/2}\tau(w)$. Since this family of distributions possesses increasing monotone likelihood ratio in W; with ϵ being the parameter; we will have $A(y)\le 0$ whenever

$$P[\frac{\partial}{\partial W}(1-y^2W)W\ge 0]=1 \text{ for all } y\in(0,2b).$$

But this occurs whenever $P[W\le\frac{1}{4b^2}]=1$, or equivalently $P[V\le\frac{1}{4b^2}]=1$.

Turning our attention now to $B(y)$, since $G''(y)\le 0$ for $y\in(0,2b)$; as above under the assumption $P(V\le\frac{1}{4b^2})=1$; we have $h''(y)\le(\frac{G'(y)}{G(y)})^2$, implying $B(y)\le(\frac{-G'(y)}{G(y)})^3-\frac{1}{y^3}$, and telling us that $B(y)\le 0$ as soon as $-y\frac{G'(y)}{G(y)}\le 1$. Finally, the result follows since

$$-y\frac{G'(y)}{G(y)}=y^2\frac{E[\,V^{3/2}\,e^{-y^2V/2}\,]}{E[\,V^{1/2}\,e^{-y^2V/2}\,]}\le 4b^2\frac{E[\,V^{3/2}\,e^{-y^2V/2}\,]}{E[\,V^{1/2}\,e^{-y^2V/2}\,]}\le 1\,,$$

for all $y \in (0, 2b)$, given the assumption $P[V \leq \frac{1}{4b^2}] = 1$.

6.3. Proof of part (c) and (d) of Lemma 7

(c) Let $\varsigma(y) = \exp\{-h(y)\}$ for all $y \in \mathbb{R}$, where h is taken from (12). Let $c, \theta > 0$ and η be given implicitly by

$$\int_{-c}^{c} (x - \eta)\varsigma(x - \theta)dx = 0.$$

Observe that $0 < \eta < \theta$ and $\eta < c$. We wish to establish that

$$\int_{-c}^{c} (x - \eta)^3 \varsigma(x - \theta)dx < 0.$$

Set

$$\psi(x) = \varsigma(|\eta - x| - (\eta - \theta)).$$

Observe that ψ is a symmetric function about η; $\psi(x) = \varsigma(x - \theta)$ for $x \leq \eta$; and $\psi(x) \leq \varsigma(x - \theta)$ for all x since $0 < \eta < \theta$. We have

$$\begin{aligned} 0 &= \int_{-c}^{c} (x - \eta)\varsigma(x - \theta)dx - \int_{2\eta - c}^{c} (x - \eta)\psi(x)dx \\ &= \int_{\eta}^{c} (x - \eta)\{\varsigma(x - \theta) - \psi(x)\}dx + \int_{-c}^{2\eta - c} (x - \eta)\varsigma(x - \theta)dx. \end{aligned}$$

Furthermore, we have $2\eta - c < \eta$. Set

$$K = \int_{-c}^{2\eta - c} |x - \eta|\varsigma(x - \theta)dx,$$

which permits us to write

$$\int_{\eta}^{c} (x - \eta)\{\varsigma(x - \theta) - \psi(x)\}dx = K\,,$$

and

$$\int_{-c}^{2\eta - c} (x - \eta)\varsigma(x - \theta)dx = -K.$$

Finally,

$$\begin{aligned}
\int_{-c}^{c}(x-\eta)^3\varsigma(x-\theta)dx &= \int_{-c}^{c}(x-\eta)^3\varsigma(x-\theta)dx - \int_{2\eta-c}^{c}(x-\eta)^3\psi(x)dx \\
&= \int_{\eta}^{c}(x-\eta)^2[(x-\eta)\{\varsigma(x-\theta)-\psi(x)\}]dx \\
&\quad + \int_{-c}^{2\eta-c}(x-\eta)^2[(x-\eta)\varsigma(x-\theta)]dx \\
&< (c-\eta)^2K - ([2\eta-c]-\eta)^2K \\
&= 0.
\end{aligned}$$

(d) To assess the concavity of η_c on $(0,\infty)$ a straightforward differentiation for models in (12) yields the identity

$$\frac{\partial}{\partial\theta}E_\theta[X^k] = E_\theta[X^k(h'(X-\theta)-\kappa_c'(\theta))],$$

for all $\theta \in \mathbb{R}$, and $k = 0, 1, 2, \ldots$ In particular for a normal model with $h(y) = y^2/(2\sigma^2)$, $y \in \mathbb{R}$, the above identity yields for $k = 0, 1, 2$ and $\theta \in \Re$:

$$\begin{aligned}
\kappa_c'(\theta) &= \frac{1}{\sigma^2}(\eta_c(\theta)-\theta); \\
\frac{\partial}{\partial\theta}\eta_c(\theta) &= \frac{1}{\sigma^2}E_\theta[X(X-\eta_c(\theta))] = \frac{1}{\sigma^2}\mathrm{Var}_\theta(X); \\
\text{and } \frac{\partial^2}{\partial^2\theta}\eta_c(\theta) &= \frac{\partial}{\partial\theta}E_\theta(X^2) - \frac{\partial}{\partial\theta}\eta_c^2(\theta) \\
&= \frac{1}{\sigma^2}\{E_\theta[X^2(X-\eta_c(\theta)] - 2\eta_c(\theta)(E_\theta[X^2]-\eta_c^2(\theta))\} \\
&= \frac{1}{\sigma^2}E_\theta[(X-\eta_c(\theta))^3].
\end{aligned}$$

Finally, the concavity of η_c on $(0,\infty)$ follows by Lemma 7 part (c), thus establishing the result.